JHU vision lab

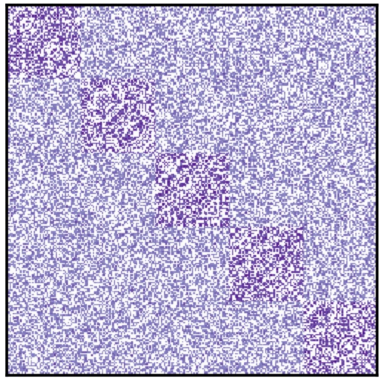
Understanding Doubly Stochastic Clustering

Tianjiao Ding[†], Derek Lim[‡], René Vidal[†], Benjamin D. Haeffele[†]

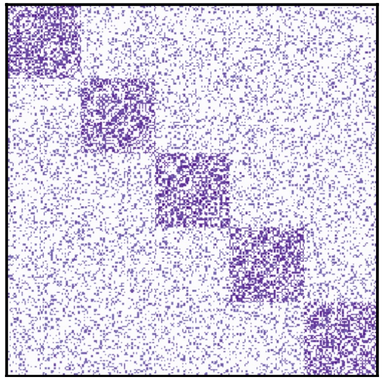
[†]Mathematical Institute for Data Science, Johns Hopkins University
[‡]Computer Science & Artificial Intelligence Laboratory, Massachusetts Institute of Technology

Spectral Clustering

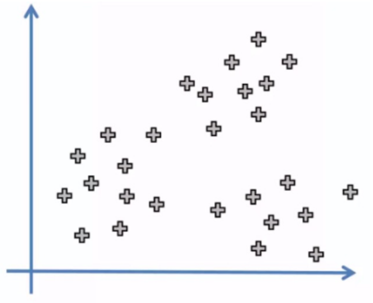
1. Define affinity matrix $\boldsymbol{K}$



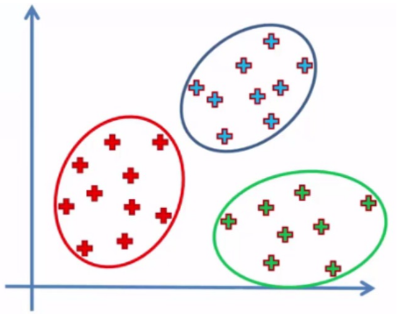
2. Compute normalized affinity $\boldsymbol{A}$



3. Compute embedding of the data from eigenvectors of Laplacian $\boldsymbol{L} = \boldsymbol{D}_{\boldsymbol{A}} - \boldsymbol{A}$

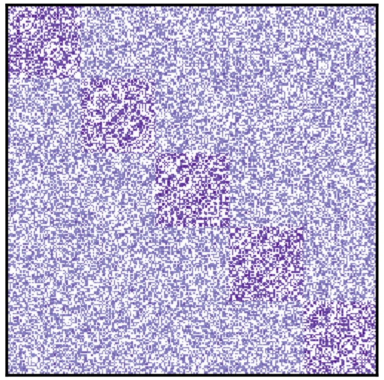


4. Cluster the embedding

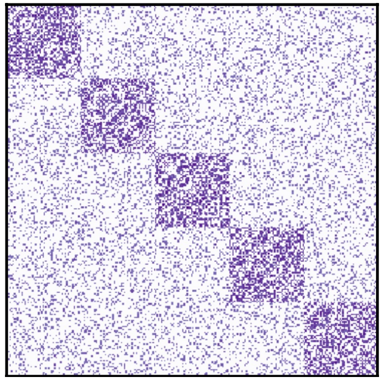


Spectral Clustering

1. Define affinity matrix $\boldsymbol{K}$



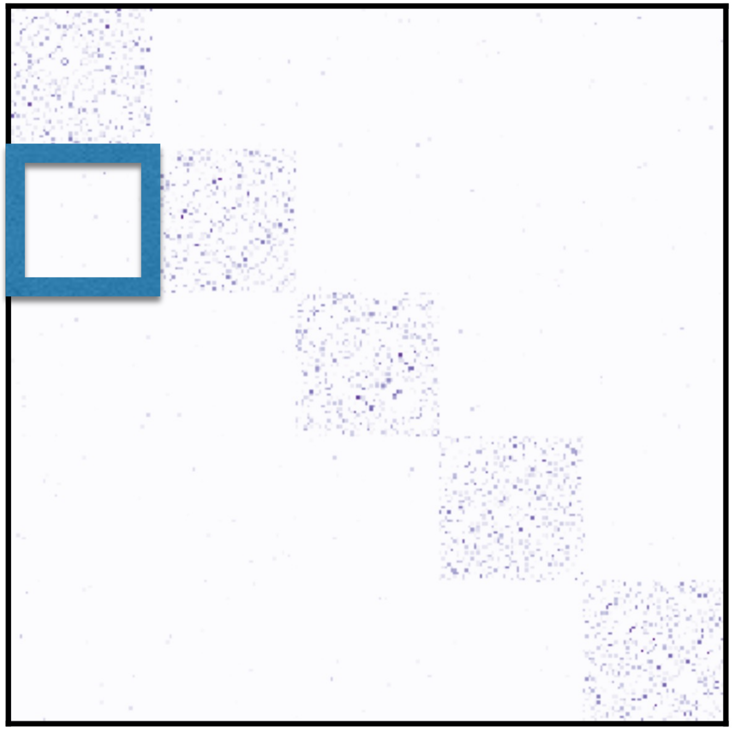
2. Compute normalized affinity $\boldsymbol{A}$

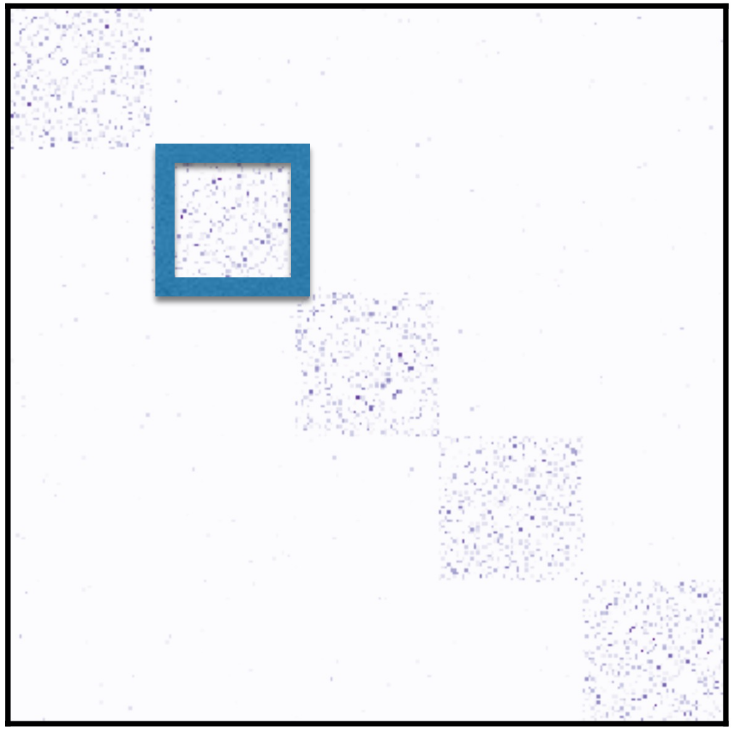


- Clustering performance depends on affinity quality
- How to define/normalize affinity?

Ideal Affinity for Clustering

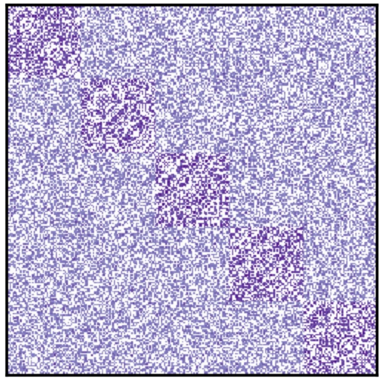
- No False Connections (NFC) for points between different clusters
- Connectivity: points within the same cluster are well connected





How to Define a Good Affinity for Spectral Clustering?

Define affinity matrix $\boldsymbol{K}$



- Design affinity based on data geometry
  - Clusters are centroids
  - Clusters are linear subspaces [1-4]

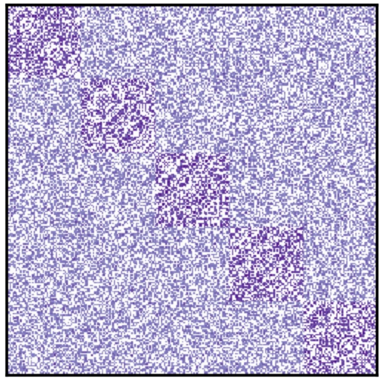
Most works only guarantee NFC but not connectivity

We give guarantees on both NFC and connectivity

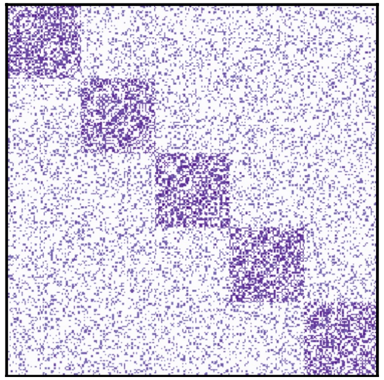
[1] E. Elhamifar and R. Vidal, Sparse Subspace Clustering, 2009.
[2] G. Liu et al, Robust Subspace Segmentation by Low-Rank Representation, 2010.
[3] C. Lu et al, Robust and Efficient Subspace Segmentation via Least Squares Regression, 2012.
[4] C. You et al, Oracle based active set algorithm for scalable elastic net subspace clustering, 2016.

How to Define a Good Affinity for Spectral Clustering?

Define affinity matrix $\boldsymbol{K}$



Normalized $\boldsymbol{A}$ by thresholding $\boldsymbol{K}$

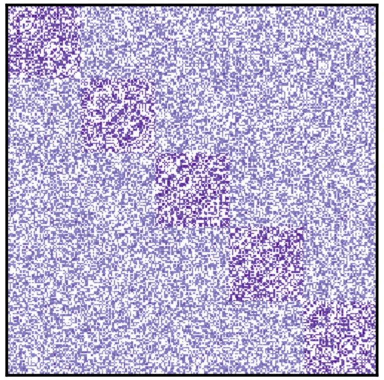


- Use normalization to clean up affinity?
  - ❌ Sparsify the affinity by thresholding

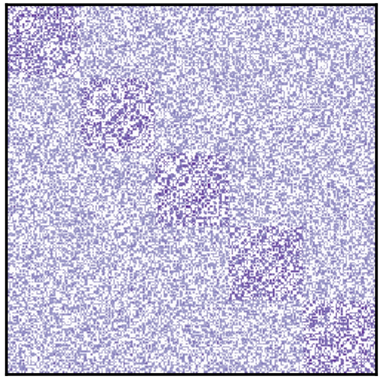
[1] Shi, J. and Malik, J. Normalized cuts and image segmentation, 2000.
[2] Zass, and Shashua, Doubly Stochastic Normalization for Spectral Clustering, 2006.
[3] Lim et al, Doubly stochastic subspace clustering, 2020.

How to Define a Good Affinity for Spectral Clustering?

Define affinity matrix $\boldsymbol{K}$



Symmetric normalized affinity $\boldsymbol{A}$



- Use normalization to clean up affinity?
  - ❌ Sparsify the affinity by thresholding
  - ❌ Normalized cut [1]

[1] Shi, J. and Malik, J. Normalized cuts and image segmentation, 2000.
[2] Zass, and Shashua, Doubly Stochastic Normalization for Spectral Clustering, 2006.
[3] Lim et al, Doubly stochastic subspace clustering, 2020.

How to Define a Good Affinity for Spectral Clustering?

Define affinity matrix $\boldsymbol{K}$

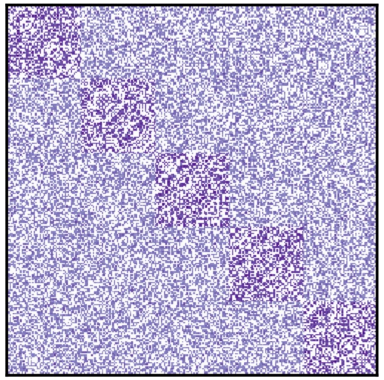
Doubly stochastic normalized $\boldsymbol{A}$

- Use normalization to clean up affinity?
  - ❌ Sparsify the affinity by thresholding
  - ❌ Normalized cut [1]
  - ✅ Doubly stochastic projection [2,3]

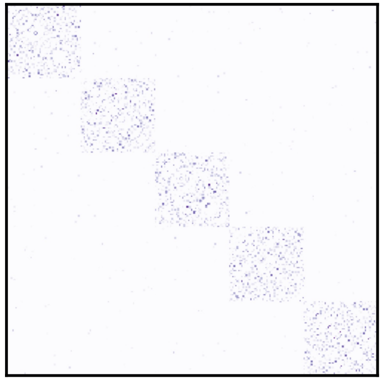
[1] Shi, J. and Malik, J. Normalized cuts and image segmentation, 2000.
[2] Zass, and Shashua, Doubly Stochastic Normalization for Spectral Clustering, 2006.
[3] Lim et al, Doubly stochastic subspace clustering, 2020.

Why Doubly Stochastic Projection?

Define affinity matrix $\boldsymbol{K}$



Doubly stochastic normalized $\boldsymbol{A}$



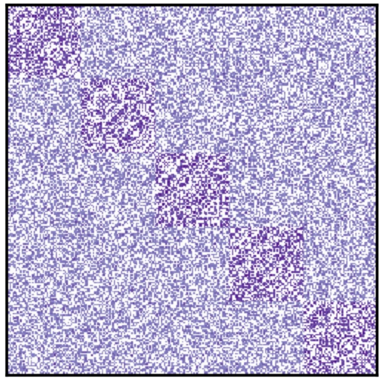
- $\mathcal{A}$: the set of doubly stochastic matrices
  - Normalized cut $\approx$ the closest matrix in $\mathcal{A}$ to $\boldsymbol{K}$ under KL divergence [1]
  - Doubly stochastic projection $:=$ the closest matrix in $\mathcal{A}$ to $\boldsymbol{K}$ under ℓ_2 metric
- Doubly stochastic projection achieves SOTA clustering performance [2]
  - E.g., 98.4% clustering accuracy on COIL-100, 99% on MNIST

Theoretical understanding? This paper: ✅

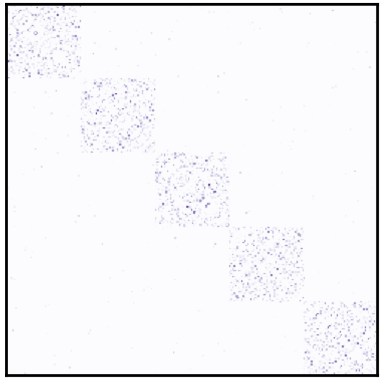
[1] Zass, and Shashua, Doubly Stochastic Normalization for Spectral Clustering, 2006.
[2] Lim et al, Doubly stochastic subspace clustering, 2020.

Contributions

Define affinity matrix $\boldsymbol{K}$



Doubly stochastic normalized $\boldsymbol{A}$



- ✅ Provable guarantees for $\boldsymbol{A}$ to have no false connections and be well-connected
- ✅ Additional guarantees for subspace clustering (each cluster is a subspace)
  - Guarantees depend on interpretable quantities (angles between subspaces, etc.)

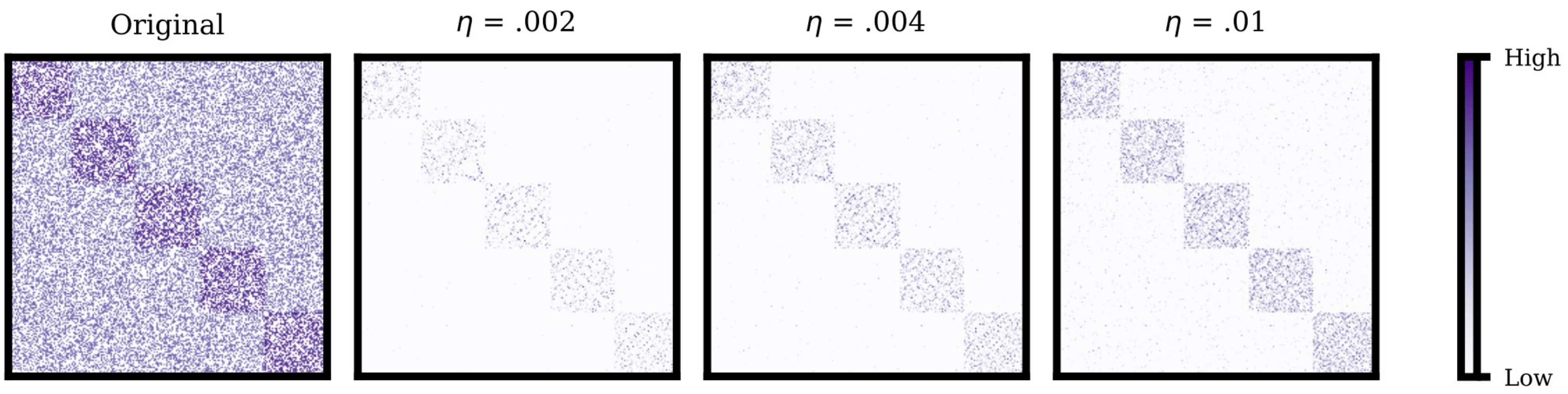
Doubly Stochastic Normalization

- Given input affinity $\boldsymbol{K} = \boldsymbol{K}^\top \in \mathbb{R}^{n\times n}$, doubly stochastic normalized $\mathbf{A}^*$ is given by

$$\mathbf{A}^* = \underset{\boldsymbol{A}\in\mathcal{A}}{\operatorname{argmin}} \left\| \boldsymbol{A} - \frac{1}{\eta}\boldsymbol{K} \right\|_{\boldsymbol{F}}$$

set of doubly stochastic matrices

$\eta > 0$: parameter such that $\mathbf{A}^*$ is sparser as $\eta \downarrow$



✅ Compute the spectral embedding directly from $\mathbf{A}^*$

Theorem: Optimality Conditions

$$\min_{\boldsymbol{A}\in\mathcal{A}} \left\| \boldsymbol{A} - \frac{1}{\eta}\boldsymbol{K} \right\|_F$$

$$\boldsymbol{\alpha}^* \in \mathbb{R}^n: \quad \frac{1}{n}\sum_{j=1}^{n} \max(K_{ij} - \alpha_i^* - \alpha_j^*, 0) = \eta, \forall i \qquad (\#)$$

$$A_{ij}^* = \frac{1}{\eta}\mathrm{ReLU}(K_{ij} - \alpha_i^* - \alpha_j^*)$$

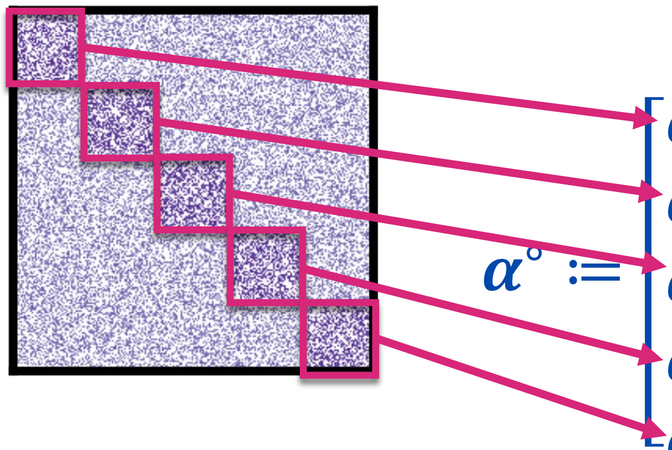
- $\boldsymbol{A}^*$ has no false connections $\Leftrightarrow K_{ij} \le \alpha_i^* + \alpha_j^*$ for inter cluster connections i, j
- $\boldsymbol{A}^*$ is well connected $\Leftrightarrow K_{ij} > \alpha_i^* + \alpha_j^*$ for intra cluster connections i, j

- Problem: Hard to bound $\boldsymbol{\alpha}^*$ due to (#) coupling among different clusters

Theorem: Decoupling is equivalent to NFC

Given input affinity $\boldsymbol{K} = \boldsymbol{K}^\top \in \mathbb{R}^{n \times n}$ and sparsity parameter η,

- $\boldsymbol{A}^*$ has no false connections $\Leftrightarrow$



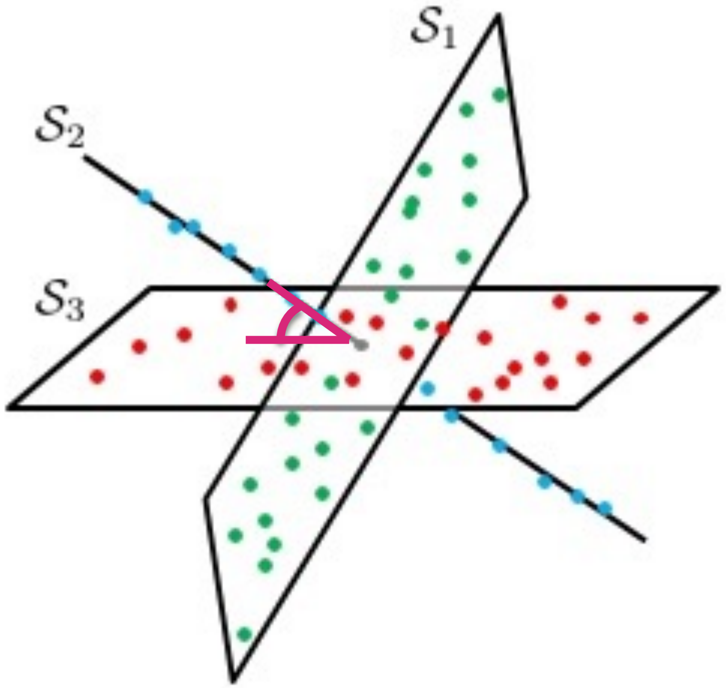
$$\boldsymbol{\alpha}^\circ \coloneqq \begin{bmatrix} \alpha^{(1)} \\ \alpha^{(2)} \\ \alpha^{(3)} \\ \alpha^{(4)} \\ \alpha^{(5)} \end{bmatrix} = \boldsymbol{\alpha}^*$$

- $\boldsymbol{\alpha}^{(l)}$ is a solution to (#) of the submatrix of intra-cluster connections of cluster l

✅ Bounding $\boldsymbol{\alpha}^\circ$ is much easier

Theorem: Subspace Clustering (Informal)

- Subspace clustering
  - clustering data from a union of low-dimensional linear subspaces
  - each subspace defines a cluster



✅ We prove: $\boldsymbol{A}^*$ has no false connection if the subspaces
  - are sufficiently separated in angle, or
  - have sufficiently low dimensions, or
  - are well balanced in terms of number of points.

✅ We also guarantee connectivity!

More Information

Research supported by NSF grant 1704458 and Northrop Grumman Mission Systems Research in Applications for Learning Machines (REALM) initiative.

Vision Lab @ JHU
http://www.vision.jhu.edu

Center for Imaging Science @ JHU
http://www.cis.jhu.edu

Mathematical Institute for Data Science @ JHU
http://www.minds.jhu.edu

Thank You!